(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷: **G06F 19/00**

(21) International Application Number:
PCT/EP2004/005368

(22) International Filing Date: 19 May 2004 (19.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0301458-6   20 May 2003 (20.05.2003)   SE

(71) Applicants and
(72) Inventors: **LINDSTRÖM, Lars** [SE/SE]; Lekevallsgatan
46, S-431 69 Mölndal (SE). **LUNDHOLM, Kent** [SE/SE];
Birkavägen 7, S-429 33 Kullavik (SE).

(74) Agent: **AWAPATENT AB**; P.O. Box 11394, S-402 28
Göteborg (SE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
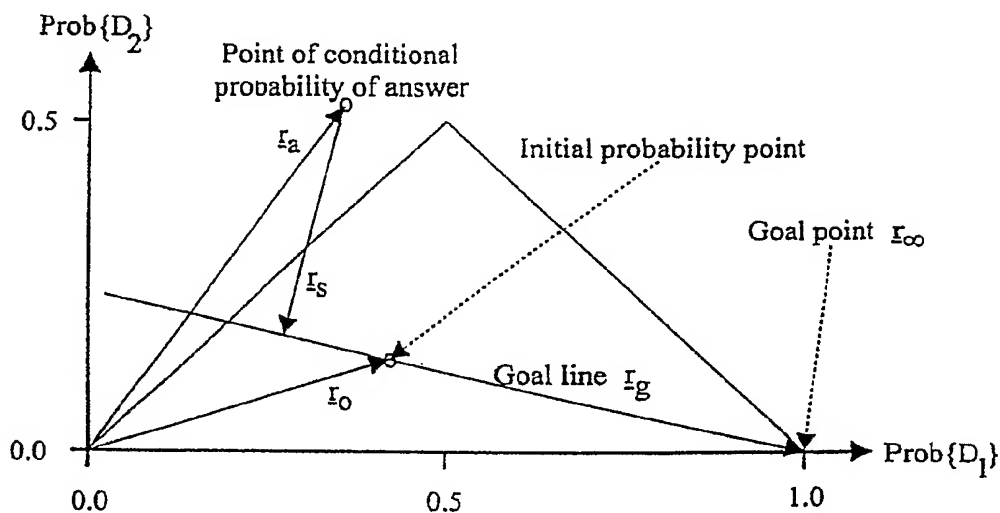TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished
upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: METHOD FOR AUTOMATED DIFFERENTIAL EVALUATION

(57) **Abstract:** A method for automated differential evaluation is disclosed, to be used, for example, for medical diagnosis or any
other diagnostic procedure based on probabilities, such as technical fault detection or any other choice situation. The method com-
prises the steps of: a) providing based on available data at least two hypotheses, to each of which is assigned a certain probability to
be valid, and selecting as main hypothesis the one of said at least two hypotheses having the highest probability, b) defining a goal
line in the probability space, between the goal point and the current probability situation, c) choosing from at least two sets of feasible
supplementary data, the one for which the effect of said supplementary data as a projection on said goal line is the greatest, and d)
reevaluating said main hypothesis based on said chosen set of supplementary data and on the experimental response thereupon.

1

METHOD AND APPARATUS FOR AUTOMATED DIFFERENTIAL
EVALUATION

Field of the Invention
The present invention relates to a method and an
apparatus for automated differential evaluation, e.g. so
5   called preferred choice evaluation, and more specifically
to a method and system for automated differential
evaluation related to medical and clinical purposes.

Background of the Invention
10   Previous attempts to improve health care related
diagnoses have involved automation in various ways, e.g.
dial-in libraries of answers to medical questions. Other
ways have involved computerized aids for patient
examination.
15   WO 98/02836, for example, shows an automated system
for clinical diagnosis. The system comprises lists
connecting various states of illness to different types
of symptoms. From the lists are generated questions to be
posed to a patient in a dialogue with the patient so that
20   information may be brought forward about the patient's
symptoms. The symptoms are weighted differently, given
different states of illness. The diagnosis starts out
from the symptom perceived as most serious by the patient
and the states of illness for which the particular
25   symptom is included are chosen. A hypothesis is then
identified from the probability of the symptom to occur.
A problem with this kind of ordinary hypotheses
testing is that only the ratio between the a posteriori
probabilities is considered for different investigations
30   without considering the final goal (verification of the
main hypothesis). Such procedures give rise to a large
number of possible combinations of a priori probabilities
of different diagnoses and investigation results.

SUBSTITUTE SHEET (RULE 26)

2

In another document, US 5935060, is shown a system and a method for providing computerized, knowledge-based medical diagnostic advice. In this case, the medical advice is provided to the user over a network, such as a
5   telephone network, or, alternatively, the advice may be provided to a patient in a stand-alone mode by use of a computer. The invention uses a list-based processing method of generating and executing diagnostic scripts. Medical knowledge is organized into a list of diseases to
10   be considered, wherein each disease includes a list of symptoms. The list of symptoms is then further described as a response to a list of questions asked to a patient about the symptom. The list structure is converted by suitable data structure transformations into a script,
15   and when a patient requires diagnosis, the script is played back as a sequence of questions. The responses are analysed and converted into symptoms, which, in turn, are accumulated into diseases. Finally, diseases are selected and reported as a diagnosis.
20   A problem with the approach in US 5935060 (as well as in WO 98/02836, for example) is that the procedure does not allow for choosing questions in an optimal way. Rather, in these cases scripts are generated and read one by one, meaning that the diagnoses cannot be built on the
25   whole picture, with the risk of leading to inadequate medical advice.
Furthermore, there is the risk that examinations and diagnoses become unnecessarily complicated and expensive when the answers to the questions do not gradually
30   support the probability for a certain diagnosis. Any unnecessary extra verification step inevitably leads to expensive and timeconsuming tests and examinations. That is to say, prior art methods does not use the a priori probabilities of the diagnoses under consideration as a
35   total appraisal of all investigations on a patient. This causes an unnecessarily complex multidimensionality of the diagnostic procedure, since all earlier details of

3

investigational results have to be considered in detail
in order to put forward a diagnosis.

Altogether this may lead to cumbersome, time-
consuming and costly processes, which are likely to cause
5   much discomfort for the patient as well as the examiner.
Furthermore, there is the risk that these processes
through their inherent complexity lead in the wrong
direction, which may reduce the reliability of the
diagnoses.

10   Similar problems are experienced in many other
situations in which one, for example, wants to minimize
the search for errors in systems with complex
interrelationships and interdependencies, other than the
human body, such as fault detection for technical
15   devices, e.g. in cars, or searching for disturbances in
process industry plants, or communication errors in large
networks, or disturbances in traffic systems. In all
these examples, the errors need to be quickly diagnosed
and corrected.

20   There is therefore a need for an improved method and
an apparatus for automated diagnostic advice.

Also, the use of decision support systems based on
rules (executing scripts or following decision trees) can
lead to large amounts of extra tests and investigations
25   in order to correct an erroneous diagnostic path caused
by atypical results of a test or investigation. The
present method prevents such errors to occur, and thus
avoids extra diagnostic costs (and uncomfort for the
patient).

30

Summary of the Invention
In a medical environment, uncertainty is a common
feature. Uncertainty makes it difficult to use a system
built on a logic with statements that are either true or
35   false, since it does not deal with uncertainty in an
effective way. To use only the current diagnostic
criteria for differential diagnosis of various diseases

4

would be at the risk of having a significant number of
the diagnoses be inadequate and diseases would
consequentially pass undetected. For the situation when a
patient is not a "textbook case", there has to be a
differential diagnostic procedure that does not over-look
rare but important features, and also is able to deal
with incomplete information as well as meet information
and criteria for more than one diagnosis.

It is therefore an object of the present invention to
provide a method and an apparatus for automated
differential evaluation, e.g. so-called preferred choice
evaluation, that mitigates the limitations of the prior
art described above, especially so for medical and
clinical purposes.

This object is achieved by means of a method and an
apparatus according to the appended claims.

According to a first aspect of the present invention,
it relates to a method for automated differential
evaluation, to be used, for example, for medical
diagnosis or technical fault detection, the method
comprising the steps of:

a) providing based on available data at least two
hypotheses, to each of which is assigned a certain
probability to be valid, and selecting as main hypothesis
the one of said at least two hypotheses having the
highest probability,

b) defining a goal line in the probability space,
between the goal point and the current probability
situation,

c) choosing from at least two sets of feasible
supplementary data, the one for which the dispersion of
said supplementary data as projections on said goal line
is the greatest, and

d) reevaluating said main hypothesis based on said
chosen set of supplementary data and on the experimental
response thereupon.

By way of example, to enhance the clarity of the
invention, are first in the following some aspects of the
essence of the present invention described. Assume a
general situation where a set of diseases is assigned a
5   set of corresponding probabilities $P_1...P_n$. By way of
example, this is discussed below for the specific
situation comprising two diseases, each being assigned a
corresponding probability, P_2 and P_1, respectively, where

10   P_1 is the probability for the main hypothesis, and
         P_2 is the probability for the alternative hypothesis,
so that
         $P_2+P_1 \le 1$ (which may be plotted in a triangular-shaped
probability space, as in figures 3-5).
15

The purpose of a diagnostic procedure method is to
ascertain the diagnosis so that the correct treatment can
be put in place. It is desirable that the diagnostic
procedure is as smooth as possible, meaning for example
20   that it shall not be time-consuming or expensive because
of too many questions asked that do not strengthen the
correct (main) hypothesis.
         In essence, this means that as long as $P_2<P_1$ for every
new question everything is fine, meaning that the correct
25   (main) hypothesis is strengthened. Otherwise, a re-
ranking is necessary, so that the present alternative
hypothesis from now on will be the new main hypothesis.
         With reference to figures 3-5, the angle or tilting
of a line from origo to (P_1,P_2) is a measure of the
30   current quality of the diagnostic question. Hence, it is
desired that for each new question this angle or tilting
is getting smaller and smaller, meaning, in other words,
that the line is approaching the horizontal x-axis.
         What is described above is known from prior art
35   methods. There is a problem with this, though, namely
that there is a risk that the probability P_1 has been

6

lowered even if the line described above actually is coming closer to the horizontal x-axis.

For this reason, within the present invention is introduced the concepts of goal line and goal point. The
5   goal point is where the desired goal is. Usually, it is considered ideal or desired that the goal point correspond to the point where $P_1=1$, i.e. 100 percent probability for the main hypothesis and zero probability for the alternative hypothesis. The goal line is the
10   straight line that connects the ideal goal point with the current probability situation, i.e. (P_1, P_2) for the present situation discussed.

By means of of the present invention every next question is tested before it is asked in comparison to
15   alternative questions, to see how well it supports the goal of getting a better and better diagnosis for each new question.

In principle this is done through choosing from at least two sets of feasible supplementary data, the one
20   for which the dispersion of said supplementary data as projections on said goal line is the greatest. In practice this means that perpendicular projections of different probabilities on the goal line are compared, where the best effect, corresponding to the question
25   which most quickly takes the user closer to a more probable diagnosis, is given for the one question that gives the largest spread/distance from the current probability situation along the goal line. Moreover, this means a simultaneous falsification of alternative
30   hypotheses. As indicated above, recalculations of probabilities are made question by question after each new answer.

Furthermore, the method makes it possible to take into account a true evidence-based behaviour and certain
35   utility aspects, for example situations including circumstances where it may always be better to act than

7

not doing so, even if the user or a physician is not exactly sure of the correct diagnosis.

Furthermore, within the present invention it is possible to take into account not only information availability but also other determinants such as time and cost aspects.

A preferable output from the use of the present invention, which can help the user to put the correct diagnosis, is, for example, a bar chart showing the most possible diagnoses.

The method is based on a truly scientific procedure concerning verification and falsification of hypotheses (diagnoses, etc), whereas in ordinary hypotheses testing, only the ratio between the a *posteriori* probabilities is considered for different investigations without considering the final goal (verification of the main hypothesis). Such commonly used procedures can give rise to a large number of possible combinations of a *priori* probabilities of different diagnoses and investigation results.

As stated above, the reevaluation of said main hypothesis based on at least one set of supplementary data is performed by perpendicularly projecting the supplementary data on a goal line, aiming at providing an optimal path. Normally, in methods according to the prior art, one instead looks at the quotient between the probabilities for the main hypothesis and the other hypothesis, which do not provide an optimal path. The projection of probabilities according to the present invention will be described in more detail later.

The present method, on the contrary, uses the a *priori* probabilities of the diagnoses under consideration as a total appraisal of all investigations on a patient. This implies that the multidimensionality of the diagnostic procedure is reduced since all earlier details of investigational results no longer have to be

8

considered in detail, but are contained in the *a priori*
and *a posteriori* probabilities.

Consequently, the present invention provides great
advantages compared to the prior art concerning the

5   possibility to truly optimize the diagnostic process, by
arriving at a correct diagnosis at a quicker time. This
makes the examination process cheaper and it becomes
possible to initiate correct treatments at an earlier
stage.

10 In addition, in the clinical practice the
differential diagnostic algorithm makes possible the
construction of efficient information handling systems
since the optimal selection of investigations (or
questions) automatically provides a well structured data

15 acquisition function. This acquisition system only deals
with investigations (questions, variables) that are the
most relevant for the current situation of the patient.
Such an acquisition system can also be used to prepare
for coming investigations if these are relevant and not

20 strongly dependent on possible changes of probabilities
during the next couple of questions (investigations).
This means that several time consuming events can be
initiated in advance (such as for pre-diagnostic
purposes), which decreases the total waiting time for

25 patients or processes and increases the efficiency of the
clinical work and procedures.

The method may also be used for training purposes and
as part of the examination of medical students. This may
be, for example, in the form of so called computer aided

30 educational systems, which may be adapted such that the
system instead of providing the correct diagnosis
directly asks controlling questions during a specific
examination case, the case being real or an exemplifying
case. Other embodiments of the present invention may also

35 be possible for training and educational purposes.

According to a preferred embodiment of the method,
the step of reevaluating said main hypothesis comprises

9

selecting information for said at least one set of
supplementary data providing maximum support of the main
hypothesis and/or providing maximum weakening of the
other hypotheses.

5   The investigation which most strongly increases the
possibility that a certain hypothesis is true does in one
preferred embodiment of the present invention represent
the investigation which most strongly lowers the
probability in case of an alternative result or answer of

10   the investigation. Alternatively, in another embodiment,
the present invention takes care of the situations where
the investigation which most strongly increases the
possibility that a certain hypothesis is true may not
represent the investigation which most strongly lowers

15   the probability in case of an alternative result or
answer of the investigation. For example, if a patient
experiences stomach pain, it might be that the pain is
due to both gastric ulcer and/or an attack of
appendicitis and/or grippes at the same time, possibly

20   with different seriousness, but all symptoms must be
diagnosed and treated.

All in all, however, the present method combines in
the most effective way the verification (or
falsification) of the main hypothesis with falsification

25   (or strengthening) of the alternative hypotheses. This
implies that a certain level of probability that the main
hypothesis is correct, is arrived at through a minimum
number of investigations. This also minimizes the cost
and time consumption for the diagnostic procedure.

30   According to a preferred embodiment of the method, it
comprises the additional step of:

(e) repeating steps (a) to (d) until a predetermined
condition is met.

The method for automated differential evaluation is

35   preferably based on user initiated questions or
automatically initiated questions, comprising the
additional step of

10

(f) choosing said main hypothesis, for which said predetermined condition is met, as the output of the differential evaluation.

The step of reevaluating said main hypothesis preferably comprises reassessing the probabilities for the hypotheses given said at least one set of supplementary data and on the experimental response thereupon.

The predetermined condition could, for example, be at least one of

- a specified number of repetitive steps being met,
- a certain probability for the main hypothesis being reached or exceeded,
- the probability for a hypothesis other than the main hypothesis reaches or falls below a certain probability,
- the probability for all hypotheses other than the main hypothesis reaches or falls below a certain probability,
- a certain policy requirement related to the evaluation being met, and
- combinations of two or more of these conditions.

Said policy requirement could be, for example, but not limited to, at least one of

- a restriction of further continuing a diagnosis due to, for example, a predetermined cost limit,
- a restriction related to money spent on diagnosis so far,
- a restriction related to a cost limit for each additional diagnostic step to be performed,
- a time constraint, e.g. in that an evaluation could be required to last no longer than a certain predetermined amount of time,
- a "utility" routine decision that the further investigation will not lead to a change in planned treatment, and

11

- a restriction caused by the patient's ability and
status.

Furthermore, said supplementary data is preferably
related to physical or clinical data. Physical data may
5   here be related to, for example, a car or other product
or process and, hence, not only to a patient. Clinical
data includes medical data, psychological data and other
data having influence on the diagnostic outcome.

In a preferred embodiment, the method, further,
10   comprises the additional step of:
- acquiring said set of supplementary data using
means for measuring data.

According to another aspect of the invention, it
relates to an apparatus for automated differential
15   evaluation, to be used, for example, for diagnosis and
fault detection, the apparatus comprising
- means for providing based on available data at
least two hypotheses, to each of which is assigned a
certain probability to be valid, and selecting as main
20   hypothesis the one of said at least two hypotheses having
the highest probability,
- means for defining a goal line in the probability
space, between the goal point and the current probability
situation,
25   - choosing from at least two sets of feasible
supplementary data, the one for which the dispersion of
said supplementary data as projections on said goal line
is the greatest, and
- means for reevaluating said main hypothesis based
30   on said chosen set of supplementary data and on the
experimental response thereopon.

As for the apparatus according to the present
invention, the advantages obtained correspond to those of
the method according to the present invention, discussed
35   above.

According to another aspect of the apparatus, it
comprises means for selecting information for said at

12

least one set of supplementary data providing maximum
support of the main hypothesis and/or providing maximum
weakening of the other hypotheses.

According to another aspect of the apparatus, it
comprises

- means for repeating steps (a) to (d) until a
predetermined condition is met.

According to another aspect of the apparatus, it
comprises

- means for choosing said main hypothesis, for which
said predetermined condition is met, as the output of the
differential evaluation.

Further, the apparatus preferably comprises

- means for reassessing the probabilities for the
hypotheses given said at least one set of supplementary
data and the experimental response thereupon.

Still further, the apparatus preferably comprises

- means for acquiring said set of supplementary data,
and

- means for measuring data.

According to another aspect of the invention, it
relates to a computer software for automated differential
evaluation, to be used, for example, for medical
diagnosis or technical fault detection, comprising code
for execution of the steps:

a) providing based on available data at least two
hypotheses, to each of which is assigned a certain
probability to be valid, and selecting as main hypothesis
the one of said at least two hypotheses having the
highest probability,

b) defining a goal line in the probability space,
between the goal point and the current probability
situation,

c) choosing from at least two sets of feasible
supplementary data, the one for which the dispersion of
said supplementary data as projections on said goal line
is the greatest, and

13

d) reevaluating said main hypothesis based on said chosen set of supplementary data and on the experimental response thereupon.

As for the computer software according to the present invention, the advantages obtained correspond to those of the method and apparatus according to the present invention, discussed in the previous.

The computer software preferably comprises additional code for execution of the steps:

- selecting information for said at least one set of supplementary data providing maximum support of the main hypothesis and/or providing maximum weakening of the other hypotheses.

The computer software, further, preferably comprises additional code for execution of the step:

- repeating steps (a) to (d) until a predetermined condition is met.

Still further, the computer software preferably comprises additional code for execution of the step:

- choosing said main hypothesis, for which said predetermined condition is met, as the output of the differential evaluation.

The computer software also preferably comprises additional code for execution of the step:

- reassessing the probabilities for the hypotheses given said at least one set of supplementary data and the experimental response thereupon.

Based on what has been said above, it is evident that the present invention mitigates the limitations as described above and, hence, it represents a significant advance in the art.

Further advantages of the present invention will, in the following, be more apparent to those skilled in the art.

14

Brief description of the drawings

The features of the present invention will be more
apparent from the following detailed description of the
invention, and reference to the drawings, wherein:

5   Fig. 1 is a schematic block diagram of a system
according to one embodiment of the present invention,

Fig. 2 is a schematic flow chart illustrating the
method according to one embodiment of the present
invention,

10   Fig. 3 is an illustration of the probability space
for the case of two competing diagnoses,

Fig. 4 is an illustration of the projection of the
conditional probability vector for the case of two
diagnoses. Symbols are: $\underline{r}_0$ is the a priori probability

15   vector, $\underline{r}_a$ is the conditional probability vector for one
of the answers to a question (or investigation), $\underline{r}_s$ is
the projection of $\underline{r}_a$ on the goal line vector $\underline{r}_g$, which
points to the ultimate diagnostic result.

Fig. 5 is an illustration of the projection of the

20   conditional probabilities of - in this case - three
possible answers to a particular question (or
investigation) for the case of two diagnoses. All
conditional probabilities of the answers are projected on
the goal line $\underline{r}_g$ and a measure of goodness is calculated

25   from the spread of the projected points.

Detailed description of the invention

The invention will now, by way of example, be
described in more detail, with reference to the drawings.

30   According to the present invention a method,
apparatus and computer software for automated
differential evaluation is provided, wherein the
evaluation procedure is made efficient by an optimized
selection of questions or investigations. The method is

35   based on statistical calculations of probabilities for
the diagnoses under consideration. The method implicitly
assumes that the conditional probabilities for the

answers to a question or the outcomes of an investigation
are known for each diagnosis.

These conditional probabilities are obtained from the
literature, medical records, patient interviews, or, in
5   the case of non-medical applications, from available
statistics concerning peoples preferences, traffic
counting, and many other kinds of statistical data
compilations. This makes it possible to give extra high
ranking to especially dangerous or fatal illnesses or
10   diseases.

The automated differential evaluation procedure
involves a series of questions and investigations, each
resulting in a change of probabilities for the diagnoses.

Fig. 1 is a block diagram of an apparatus and system
15   according to one embodiment of the present invention,
comprising at least one each of a database 1, a computer
2, a communication network 3, a measurement means 4, and
an interface 5.

The database 1 may, for example, contain lists, or
20   other means for systematization of data. The data could
concern symptoms, diseases, diagnoses, patient data,
drugs, treatment data, or the like, i.e. in the case of
the method and apparatus being used for medical and/or
clinical purposes. In particular, the data consist of
25   conditional probabilities linking answers to questions or
results of investigations to the various diagnoses.

If the method and apparatus is used for other
purposes, as described later, other types of data may of
course be stored in the database, such as data related to
30   fault detection in cars, or business decision related
data. Database 1 is preferably connected to a computer 2,
which could e.g. be a general purpose computer, via a
network 3 or other suitable communication means. The
network 3 maybe a wired or wireless local network, such
35   as a LAN, or it may be the Internet, for example. Stand
alone systems are also possible. Input means could be
arranged for manual input of data, e.g. by means of a

16

keyboard or a microphone on the computer 2, for entering
of data by e.g. the medical practitioner. Furthermore,
the input means could comprise measurement means 4
incorporated in the system. Said measurement means may be
5   of various kind, e.g. medical equipments, such as
electrocardiographs or any other medical equipment.
Interfaces 5 facilitates communication in the system.

Fig. 2 is a flow chart illustrating the method
according to one embodiment of the present invention,
10   which will be reviewed in the following.

In step S1, at least two hypotheses are provided
based on available data. Initially, the hypotheses are
given certain values of probability. These values can be
equal for all hypotheses (diagnoses) under consideration,
15   or they can reflect the prevalence of the diagnoses at
the specific clinical department. In cases where extra
safety is needed, diagnoses having serious outcome may be
emphasized through a higher initial probability value.

In step S2, each of these at least two hypotheses are
20   assigned a certain probability to be valid.

Thereafter, in step S3 the one of said at least two
hypotheses having the highest probability is selected as
main hypothesis.

In step S4, at least one set of supplementary data is
25   reviewed, said at least one set of supplementary data
being used to obtain experimental information.

In step S5, said main hypothesis is reevaluated based
on said review of at least one set of said supplementary
data, in combination with experimental outcome data, by
30   perpendicularly projecting the supplementary data on a
goal line, aiming at providing an optimal path.
Furthermore, a re-ranking of the probabilities for the
main hypothesis and the at least one other hypothesis is
performed.

35   In step S6, steps S2 to S5 may be repeated until a
predetermined condition is met.

17

In step S7, the correct main hypothesis is chosen as the diagnosis, i.e. the output of the differential evaluation, basically meaning that we have arrived at the correct diagnosis in an optimized way.

5   The method outlined above will now be discussed in more detail. Statistical methods and models incorporate probabilities of occurrences of, for example, diseases, while qualitative models generally use symbolic reasoning methods as "logical deduction" as in, for example, 10 combinatorial or Boolean logic. An example of a model that combines the two types is the so called Bayesian approach, which can have the capacity to capture causal, temporal and other knowledge that is hard to model in a readily way and, hence, is highly useful as 15 representation scheme for probabilistic reasoning.

The algorithm for differential diagnosis according to the present invention is used iteratively in a process based on Bayesian statistics as:

20              $$P(D|R) = P(R|D) \ P(D) \ / \ P(R) \ .$$

The notations are; $P(D|R)$ is the a *posteriori* probability that a certain diagnosis D is true after the result R of a certain investigation has been considered. 25 $P(R|D)$ is the conditional probability that we find the result R when a patient has the diagnosis D. $P(D)$ is the a *priori* probability that the diagnosis D is true prior to the investigation, and $P(R)$ is the probability that the result R will occur in the investigation. In the 30 iteration, the previous a *posteriori* probability will be used as a *priori* probability in the next step. The procedure requires knowledge of the conditional probabilities $P(R|D)$ for all diagnoses (or alternatives) under consideration and for all results of pertinent 35 investigations.

As stated earlier, these conditional probabilities are obtained from the literature, medical records, data

18

base information, patient interviews, or, in the case of non-medical applications, from available statistics concerning peoples prferences, traffic counting, and many other kinds of statistical data compilations.

5   All diagnoses or alternatives under consideration are ranked according to their *a priori* probabilities in each step, earlier referred to as "available data". In a preferred embodiment of the present invention, the differential diagnostic algorithm then selects the

10   investigation which optimally combines a strengthening of the main (highest ranked) diagnosis with falsifications of the other diagnoses or alternatives. Then, the *a posteriori* probabilities for all diagnoses under consideration are calculated and re-ranked after the

15   result of the selected investigation is known. Depending on the result of the investigation, the main diagnosis or alternative may have been strengthened or weakened or even replaced with a different one.

The diagnostic iteration (step S6) is continued until

20   further data on conditional probabilities are lacking, or the probability of the highest ranked diagnosis is considered high enough. The latter case can occur if a certain probability level is considered safe according to some policy decision, or it can occur as a result of a

25   combination of probabilities of diagnoses with a "utility" routine which decides the optimal treatment (step S7) of the patient even when high probabilities have not been achieved.

In order to avoid time consuming searches in the data

30   base 1 of conditional probabilities, one may define certain characteristic regions in the space of alternative diagnoses defined by their probabilities and calculate the optimal projections of the results of the investigations for each region. The investigations,

35   ranked after importance in the specific situation, can then be placed in look-up tables for combinations of diagnoses.

The procedure described so far have been focused on the information value of the investigations for given probabilities of medical diagnoses. Instead of the information value we may equally well use the time a certain investigation takes, or the cost it represents, or the discomfort it will cause a patient. These aspects may be treated separately and then merged in a weighted measure. Depending on the state of a patient, one of the alternative aspects may be emphasized, thus giving another set of optimal investigations. One such condition is when patients becomes critically ill, where minimal time consumption becomes important. Under less urgent conditions one can put more emphasis on cost aspects.

For the sake of clarity, we will now focus, by way of example, on a single step in the automated differential evaluation procedure according to the present invention, showing how this step can be optimized given a current panorama of probabilities, said panorama representing the combined influence of the answers to or outcomes of prior questions or investigations. It would be appreciated by someone skilled in the art that this concept of the present invention may be expanded as discussed in other places in this application, e.g. to be used in several sequential steps, to be used for several concurrent diagnoses, etc.

The diagnoses, D_n (N in number), are ranked according to their *a priori* probabilities, $\text{Prob}\{D_n\}$, with the most probable one, D_1, considered as the main hypothesis. This can be formulated as

$$\text{Prob}\{D_n\} \geq \text{Prob}\{D_{n+1}\} , \quad (n = 1 \text{ to } N-1), \tag{1}$$

with the additional condition that

$$\sum_{n=1}^{N} \text{Prob}\{D_n\} = 1. \tag{2}$$

20

This implies, for instance, that $\text{Prob}\{D_2\}$ is always less than or equal to 0.5, $\text{Prob}\{D_3\}$ is less than or equal to one-third, etc. Fig. 3 is an illustration of the probability space for the case of two diagnoses, hence the shape of the $\text{Prob}\{D_n\}$-region for two diagnoses only are shown (N=2).

In the N-dimensional probability space containing the N diagnoses under consideration, the vector $\underline{r}_0$, given by the *a priori* probabilities $\text{Prob}\{D_n\}$ (n = 1 to N), constitutes the current status of the diagnostic procedure. Under the main hypothesis that D_1 is the correct diagnosis, the final goal would be to reach the vector $\underline{r}_\infty$ characterized by unity value of $\text{Prob}\{D_1\}$ and zero probabilities for the other diagnoses, i.e., that

$$P_1(\text{goal}) = \text{Prob}\{D_1\} = 1 \tag{3a}$$

and

$$P_n(\text{goal}) = \text{Prob}\{D_n\} = 0, \quad (n = 2 \text{ to } N) \tag{3b}.$$

The shortest way to follow would then be along the line given by the two vectors $\underline{r}_0$ and $\underline{r}_\infty$. This line is denoted the goal line and is defined by the vector

$$\underline{r}_g = \underline{r}_0 + g\,\underline{u} \tag{4},$$

where g is a parameter describing the position along the goal line and $\underline{u}$ is a vector describing the orientation of the goal line and also fulfilling the relation

$$\underline{r}_\infty = \underline{r}_0 + g_\infty\,\underline{u} \tag{5},$$

where g_∞ is a parameter value associated with the final goal point.

21

The most discriminating question to be asked next is the question whose answers are associated with probabilities (for the various diagnoses) which, perpendicularly projected on said optimal goal line,

5   display the largest spread along that line. For each answer, A, to a certain question, a set of conditional probabilities for the diagnoses are obtained, $\text{Prob}\{A|D_n\}$ ($n = 1$ to N), which constitutes a vector $\underline{r}_a$ in the N-dimensional probability space. We now project the end

10   point of this vector on the goal line, i.e. we search a vector $\underline{r}_s$ from $\underline{r}_a$ which is perpendicular to the goal line $\underline{r}_g$. This is obtained with the relations

$$\underline{r}_s = \underline{r}_a + s\,\underline{m} \tag{6a},$$

15

and

$$\underline{m} \bullet \underline{u} = 0 \tag{6b},$$

20   Here, s is a parameter describing the position along the vector $\underline{r}_s$ and $\underline{m}$ is the orientation of the same vector. The "dot" in equation (6a) denotes the scalar product and expresses the conditions for perpendicularity between $\underline{r}_g$ and $\underline{r}_s$. Fig. 4 is an illustration of this

25   procedure, i.e., the projection of the conditional probability vector for the simplified case of two diagnoses.

The same projection procedure is carried out for all answers, assumed to be k in number, of the question under

30   investigation. When all projections on the goal vector are obtained, the spread of these points is calculated. The measure of spread can be the range, the standard deviation, or some other measure. If standard deviations are used instead of range, a correction for the number of

35   answers have to be made in terms of multiplication of a factor, F_c, which is

22

$$F_c = [k+1]^{\frac{1}{4}} \tag{7}.$$

Fig 5 is an illustration of the projection of the conditional probabilities of all answers to a particular question for the simplified case of two diagnoses only.

The question having the largest spread is the question whose answer potentially most strengthen the current main hypothesis and at the same time weakens the alternative hypotheses.

Instead of recalculating the projections each time an answer is obtained, certain characteristic regions of the a *priori* probabilities can be identified and their goal lines calculated. For each of these goal lines the questions can be ranked and stored in pre-calculated tables.

It should be pointed out that the present invention is not limited to the realizations described above. Similar alternative solutions are comprised by the invention, as it is defined in the claims. For example, we have in the description above referred to a case comprising two diagnoses, however, in reality the present invention allows for a large number of diagnoses.

Furthermore, the main applications of the method for automated differential evaluation, as described, are in decision support systems of medicine. However, the applications are not confined to diagnostic procedures only; also the choice of therapeutic methods can be optimized. The inventive method then allows necessary and important questions to be asked in order to suggest the best therapy. Examples are optimal drug treatment (choice of drug and administration regimen), nutrition optimization in cancer patients, etc. Accordingly, even though terms as "diagnosis", "diagnostic" and the like has been used repeatedly in the specification above, such terms should not be construed to delimit the possible areas in which the invention may find use.

23

The present invention can also be used in many other situations to minimize the search for errors in systems with complex interrelationships and dependencies. The details of these relations need not to be known; the
5   conditional probabilities of certain values of observed quantities under certain error conditions must, however, be available. Examples of applications of the algorithm for optimization of differential diagnosis for error detection are disturbances in the process industry,
10   troubleshooting in car engines, communication errors in large networks, and disturbances in traffic systems.

The method for automated differential evaluation has also applications to systems characterized by "soft" data, such as psychology and sociology. It can then be
15   used to minimize the number of questions needed to make a decision in order to achieve a certain goal. The goals can be to predict behaviour of individuals and groups of individuals, and to predict their preferences in certain questions. The application is obvious in design processes
20   of products for certain target groups, such as the number of doors in a family car, preferences for type of instruments in sports cars, etc. Likewise, the present invention is useful, for example, for solving ergonomically related design problems, such as
25   adaptations of products for physically or mentally disabled or handicapped people.

Other examples where the present invention may be used are questioning and interrogation of a suspect in a court, or a decision support system when market and
30   business decisions are to be made.

Moreover, the present invention further provides possibilities to include steps in which questions having too large correlation can be disregarded or not chosen. Within the present invention, whenever considered
35   necessary for example because of certain uncertainties experienced, it is also possible to assign various

24

supplementary data with different probability weights or "degrees of truth".

Finally, the present invention can be used with mechanical-type solutions, such as punch cards and the

5   like.

25

CLAIMS

1. A method for automated differential evaluation, to be used, for example, for medical diagnosis or technical fault detection, the method comprising the steps of:
   a) providing based on available data at least two hypotheses, to each of which is assigned a certain probability to be valid, and selecting as main hypothesis the one of said at least two hypotheses having the highest probability,
   b) defining a goal line in the probability space, between the goal point and the current probability situation,
   c) choosing from at least two sets of feasible supplementary data, the one for which the dispersion of said supplementary data as projections on said goal line is the greatest, and
   d) reevaluating said main hypothesis based on said chosen set of supplementary data and on the experimental response thereupon.

2. A method according to claim 1, wherein step (c) comprises selecting information for said at least one set of supplementary data providing maximum support of the main hypothesis and/or providing maximum weakening of the other hypotheses.

3. A method according to claim 1 or 2, comprising the additional step of:
   (e) repeating steps (a) to (d) until a predetermined condition is met.

4. A method according any one of the claims 1 to 3, wherein the automated differential evaluation is based on user initiated questions or automatically initiated questions, comprising the additional step of:

26

(f) choosing said main hypothesis, for which said predetermined condition is met, as the output of the differential evaluation.

5   5. A method according to any one of the claims 1 to 4, wherein step (d) comprises reassessing the probabilities for the hypotheses given said at least one set of supplementary data and on the experimental response thereupon.

10

6. A method according to any one of the claims 1 to 5, wherein the predetermined condition is at least one of:
      - a specified number of repetitive steps being met,
15   - a certain probability for the main hypothesis being reached or exceeded,
      - the probability for a hypothesis other than the main hypothesis reaches or falls below a certain probability,
20   - the probability for all hypotheses other than the main hypothesis reaches or falls below a certain probability,
      - a certain policy requirement related to the evaluation being met,
25   - a patient's status, and
      - combinations of two or more of these conditions.

7. A method according to any one of claims 1 to 6, wherein said supplementary data comprises physical or
30   clinical data.

8. A method according to any one of claims 1 to 7, comprising the additional step of:
      - acquiring said set of supplementary data using
35   means for measuring data.

27

9. An apparatus for automated differential evaluation, to be used, for example, for medical diagnosis or technical fault detection, the apparatus comprising:

- means for providing based on available data at least two hypotheses, to each of which is assigned a certain probability to be valid, and selecting as main hypothesis the one of said at least two hypotheses having the highest probability,

- means for defining a goal line in the probability space, between the goal point and the current probability situation,

- choosing from at least two sets of feasible supplementary data, the one for which the dispersion of said supplementary data as projections on said goal line is the greatest, and

- means for reevaluating said main hypothesis based on said chosen set of supplementary data and on the experimental response thereupon.

10. An apparatus according to claim 9, comprising:

- means for selecting information for said at least one set of supplementary data providing maximum support of the main hypothesis and/or providing maximum weakening of the other hypotheses.

11. An apparatus according to claim 9 or 10, comprising:

- means for repeating steps (a) to (d) until a predetermined condition is met.

12. An apparatus according to any one of the claims 9 to 11, comprising:

- means for (f) choosing said main hypothesis, for which said predetermined condition is met, as the output of the differential evaluation.

28

13. An apparatus according to any one of claims 9 to 12, comprising:

- means for reassessing the probabilities for the hypotheses given said at least one set of supplementary data and the experimental response thereupon.

14. An apparatus according to any one of claims 9 to 13, comprising:

- means for acquiring said set of supplementary data, and
- means for measuring data.

15. A computer software for automated differential evaluation, to be used, for example, for medical diagnosis or technical fault detection, comprising code for execution of the steps:

a) providing based on available data at least two hypotheses, to each of which is assigned a certain probability to be valid, and selecting as main hypothesis the one of said at least two hypotheses having the highest probability,

b) defining a goal line in the probability space, between the goal point and the current probability situation,

c) choosing from at least two sets of feasible supplementary data, the one for which the dispersion of said supplementary data as projections on said goal line is the greatest, and

d) reevaluating said main hypothesis based on said chosen set of supplementary data and the experimental response thereupon.

16. A computer software according to claim 15, comprising additional code for execution of the steps:

- selecting information for said at least one set of supplementary data providing maximum support of the main

29

hypothesis and/or providing maximum weakening of the other hypotheses.

17. A computer software according to claim 15 or 16, comprising additional code for execution of the steps:
    - repeating steps (a) to (d) until a predetermined condition is met.

18. A computer software according to any one of the claims 15 to 17, it comprises additional code for execution of the step:
    - choosing said main hypothesis, for which said predetermined condition is met, as the output of the differential evaluation.

19. A computer software according to any one of the claims 15 to 16, comprising additional code for execution of the step:
    - reassessing the probabilities for the hypotheses given said at least one set of supplementary data and the experimental response thereupon.
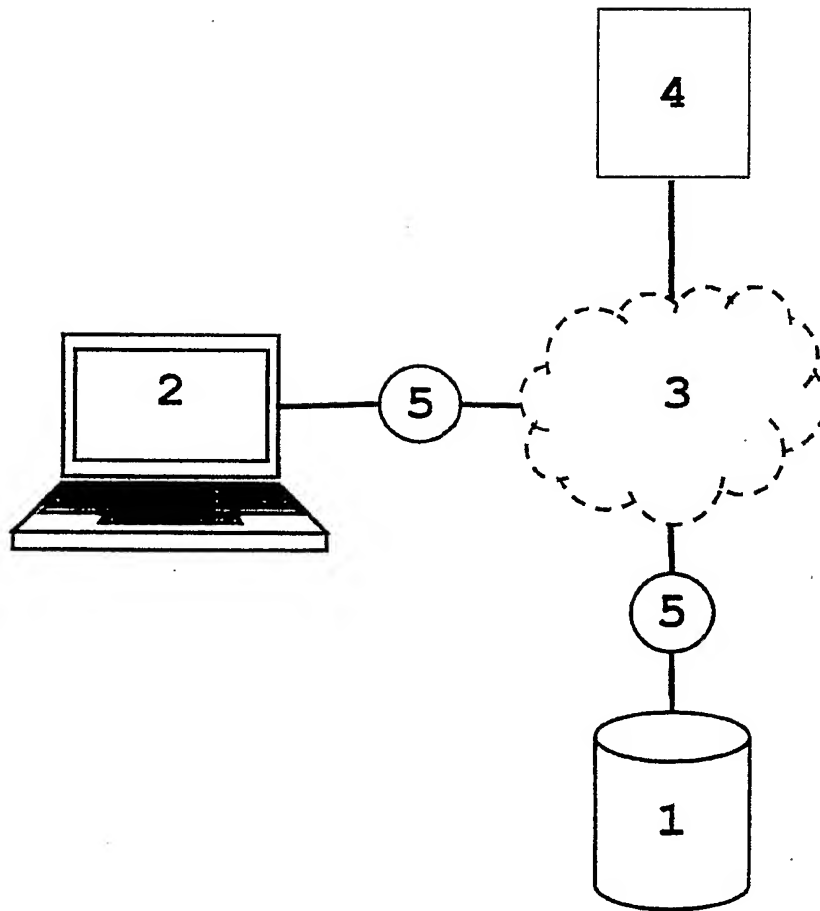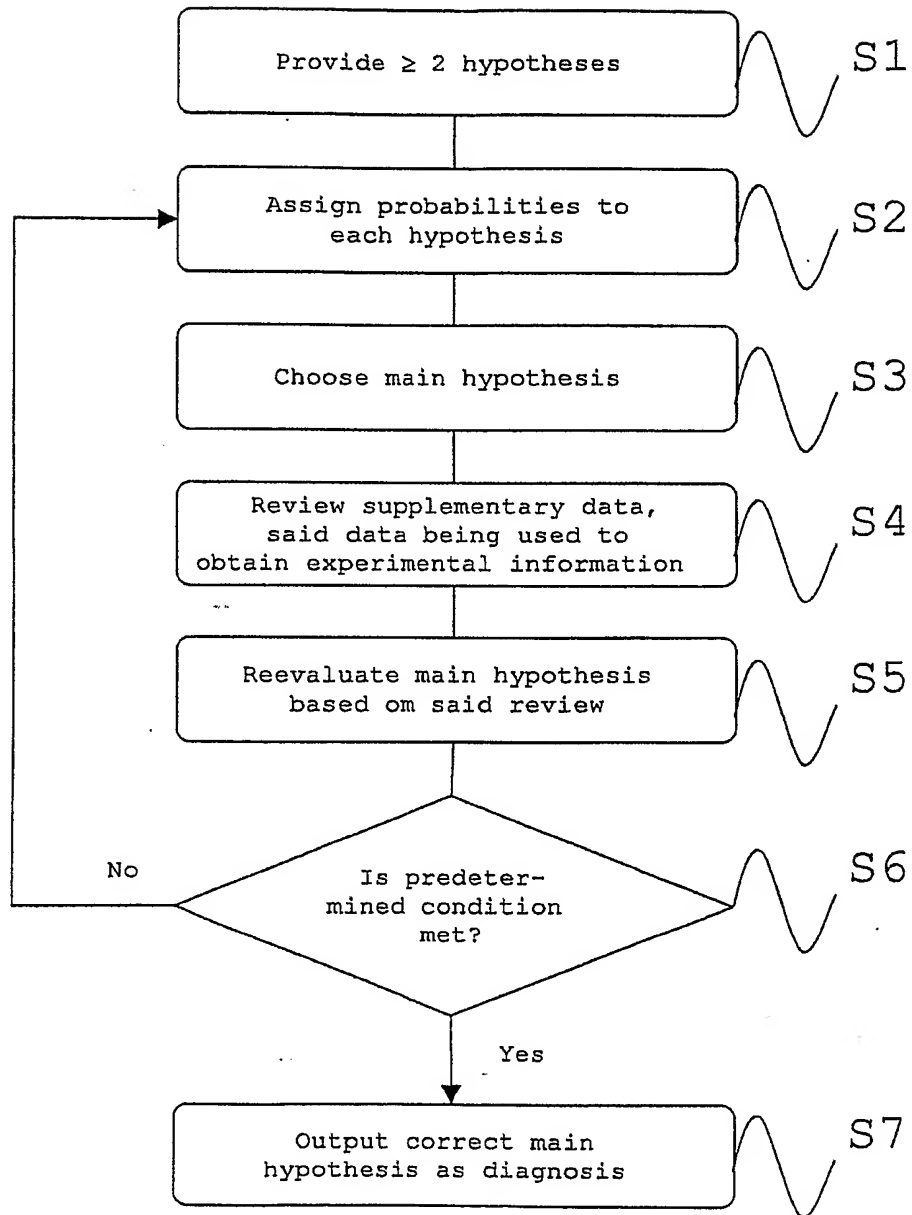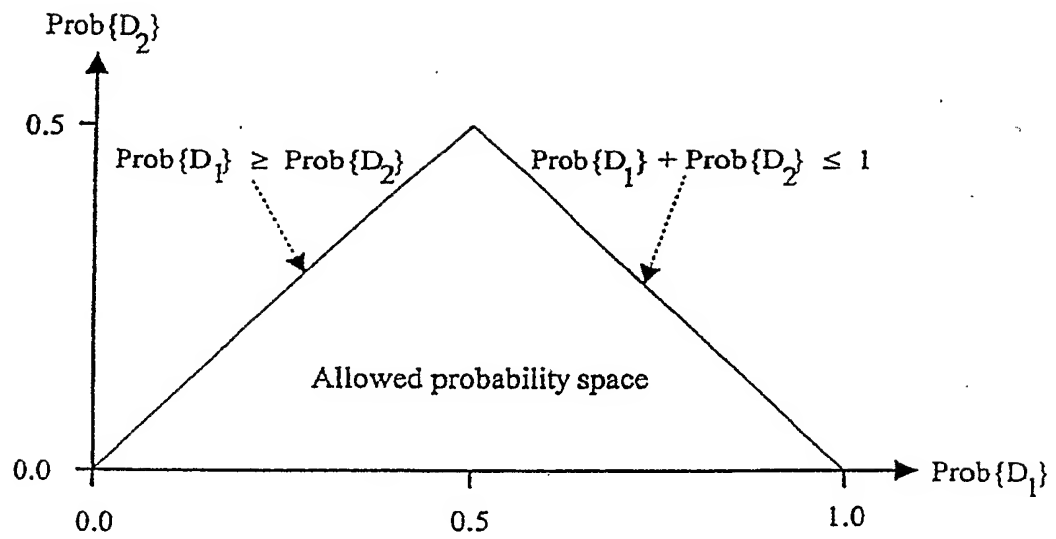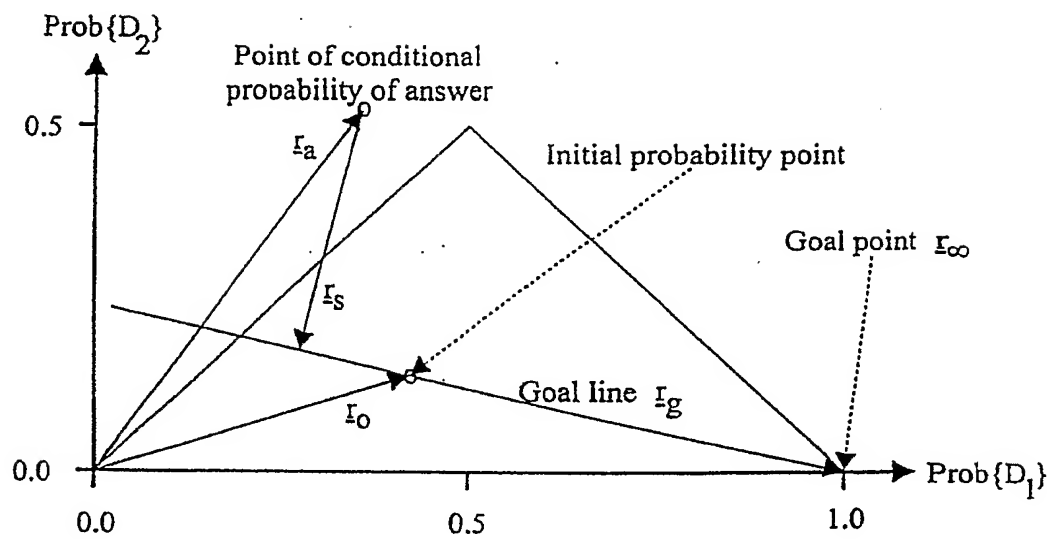
Fig. 1

Fig. 2

```
┌─────────────────────────────────┐
│      Provide ≥ 2 hypotheses      │  ∿  S1
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│      Assign probabilities to     │  ∿  S2
│         each hypothesis          │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│       Choose main hypothesis     │  ∿  S3
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│    Review supplementary data,    │  ∿  S4
│        said data being used to   │
│  obtain experimental information │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     Reevaluate main hypothesis   │  ∿  S5
│        based om said review      │
└─────────────────────────────────┘
                 │
                 ▼
          ◇─────────────◇
   No    ╱  Is predeter-  ╲          ∿  S6
 ←──────╱   mined condition ╲
        ╲       met?       ╱
         ╲─────────────────╱
                 │ Yes
                 ▼
┌─────────────────────────────────┐
│        Output correct main       │  ∿  S7
│       hypothesis as diagnosis    │
└─────────────────────────────────┘
```

Fig. 3

Fig. 4

Fig. 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷: **G06F 19/00**

(21) International Application Number:
PCT/EP2004/005368

(22) International Filing Date: 19 May 2004 (19.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0301458-6 20 May 2003 (20.05.2003) SE

(71) Applicants and
(72) Inventors: LINDSTRÖM, Lars [SE/SE]; Lekevallsgatan 46, S-431 69 Mölndal (SE). LUNDHOLM, Kent [SE/SE]; Birkavägen 7, S-429 33 Kullavik (SE).

(74) Agent: AWAPATENT AB; P.O. Box 11394, S-402 28 Göteborg (SE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,

GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
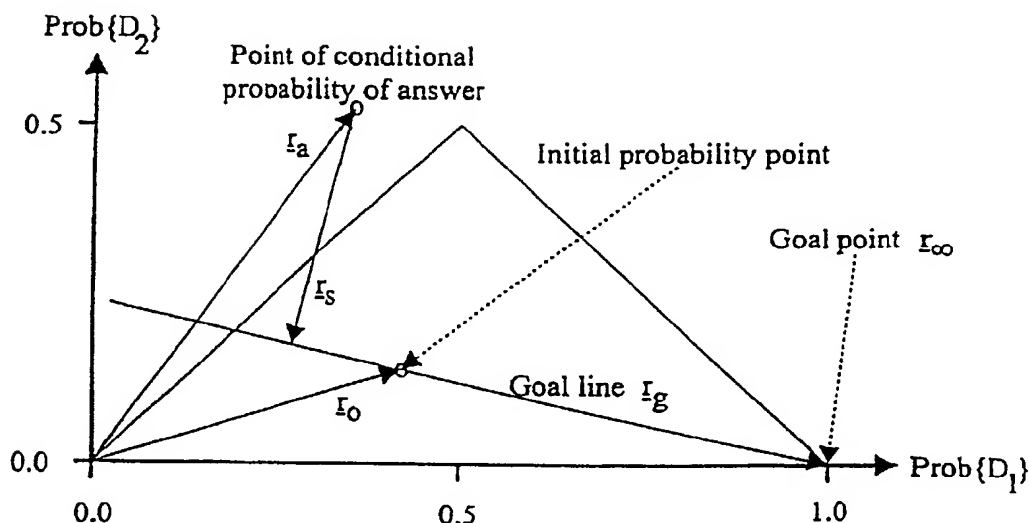
Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
28 April 2005

[Continued on next page]

(54) Title: METHOD FOR AUTOMATED DIFFERENTIAL EVALUATION

(57) Abstract: A method for automated differential evaluation is disclosed, to be used, for example, for medical diagnosis or any other diagnostic procedure based on probabilities, such as technical fault detection or any other choice situation. The method comprises the steps of: a) providing based on available data at least two hypotheses, to each of which is assigned a certain probability to be valid, and selecting as main hypothesis the one of said at least two hypotheses having the highest probability, b) defining a goal line in the probability space, between the goal point and the current probability situation, c) choosing from at least two sets of feasible supplementary data, the one for which the effect of said supplementary data as a projection on said goal line is the greatest, and d) reevaluating said main hypothesis based on said chosen set of supplementary data and on the experimental response thereupon.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC, IBM-TDB, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MCSHERRY, D: "A domain-independent theory for testing fault hypotheses" IEEE COLLOQIUIM ON 'INTELLIGENT FAULT DIAGNOSIS - PART 1 CLASSIFICATION-BASED TECHNIQUES', vol. 045, 25 February 1992 (1992-02-25), pages 1-4, XP002314881 LONDON UK 'retrieved on 2005-01-20! the whole document ----- -/--	1-19

| X | Further documents are listed in the continuation of box C. | | Patent family members are listed in annex. |

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 January 2005	15/02/2005

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Sisk, A

Form PCT/ISA/210 (second sheet) (January 2004)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	COOPER, G F: "A diagnostic method that uses causal knowledge and linear programming in the application of Bayes' formula" COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, vol. 22, 1986, pages 223-237, XP002314885 page 232, paragraph 3.5 - page 233 figure 8 figure 9 page 224, right-hand column -----	1-19
Y	HADDAWY P,KAHN JR CE, BUTARBUTAR M: "A Bayesian network model for radiological diagnosis and procedure selection: Work-up of suspected gallbladder disease" MEDICAL PHYSICS, vol. 21, no. 7, July 1994 (1994-07), pages 1185-1192, XP002314883 the whole document -----	1-19
Y	GORRY GA: "Experience with a Model of Sequential Diagnosis" COMPUTERS AND BIOMEDICAL RESEARCH, vol. 1, 1968, pages 490-507, XP002314884 the whole document -----	1-19
Y	MCSHERRY D: "Avoiding premature closure in sequential diagnosis" ARTIFICIAL INTELLIGENCE IN MEDICINE, vol. 10, 1997, pages 269-283, XP002314882 page 270 - page 271 page 275 page 280 - page 282 -----	1-19